THE JOINED-UP DATA MATURITY ASSESSMENT

ACKNOWLEDGEMENTS

The Joined-Up Data Maturity Assessment and its introductory guide were authored by Tom Orrell, Managing Director of DataReady Limited, on behalf of the Global Partnership for Sustainable Development Data, and Luis Gerardo González Morales, Statistician at the United Nations Statistics Division.

The Guide was supported by all members of the Collaborative on SDG Data Interoperability, to whom the authors extend their gratitude and thanks. The authors would like to thank in particular members of the Advisory Group from within the Collaborative who supported the production of the Maturity Assessment, including Eric Swanson at Open Data Watch; Josh Powell and Paige Kirby at Development Gateway; Karen Carter at the United Nations Children's Fund; Malarvizhi Veerappan at the World Bank; Mahsa Hedayati at the UN Office of Information and Communications Technology; and Neil Jackson at the U.K.'s Foreign, Commonwealth & Development Office.

The authors would also like to extend a special thanks to Bill Anderson at Development Initiatives, whose foresight helped put interoperability on the data revolution policy map; Shaida Badiee and the whole Open Data Watch team, who have continuously supported this work; Manuel Cuéllar and Enrique Jesus Ordaz López at Mexico's National Institute of Statistics and Geography for their continued support and feedback; Fredy Rodriguez at CEPEI (Centro de Pensamiento Estratégico Internacional); Rachael Beavan, Edafe Onerhime, and their colleagues at the Foreign, Commonwealth & Development Office; the UN Statistical Commission's Working Group on Open Data; and Francesca Perucci at the UN Statistics Division and Claire Melamed at the Global Partnership and their teams, for their support for the Collaborative on SDG Data Interoperability over the last four years.

Joined-Up Data Maturity Assessment Framework

Interoperability Governance Layers

Dimensions

ORGANIZATIONAL

	UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
Strategic Objectives	Interoperability is not recognized as a strategic objective	The ability to join up data is recognized but it is not explicitly identified as a strategic objective	Interoperability is identified as a strategic objective in an organization's technical units, but not outside of them	The need to join up data across systems is recognized as a strategic objective in an organizational data strategy The value of standards and robust data governance is recognized	The strategic value that joined-up data can bring to decision-making is recognized in organizational strategies Interoperability forms part of an organization's external engagement strategy with other data producers and users
Leadership and Management	There is no defined leadership over interoperability issues	Ad hoc leadership on interoperability issues emerges organically but is not coordinated	Leadership around interoperability emerges across various technical units but remains fragmented Silos persist	There is a coordinated hierarchy of leadership over interoperability issues Clear functions relating to interoperability are established across an organization	There is a data governance committee or council and it has an explicit mandate to lead on interoperability issues The value of joined-up data is understood by organizational leaders and managers, and is clearly identified as a function in relevant job descriptions
Oversight and Accountability	There is no oversight or accountability over interoperability issues	An organization is aware of the need to create accountability chains to oversee how data is being joined up, but is not yet taking steps to create them	Oversight structures emerge across different technical units but are not coordinated or aligned Accountability over how interoperability efforts are undertaken is fragmented and unclear	Oversight and accountability functions are embedded in an organization's strategy and reflected in the leadership structure Fragmented chains start to join up and common standard operating procedures emerge	A clear chain of oversight and accountability flows from an organization's data governance committee or council, down to operational staff Organizational units are clear about their functions relating to interoperability and who they are accountable to

ORGANIZATIONAL

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
Legal Compliance				
There is no awareness around any applicable legal obligations relating to joining up interoperable data	There is a general understanding that the actions interoperability facilitates — data transmission, sharing, and use — might be regulated, but it is unclear how	Active steps are taken to better understand legal compliance requirements around data retention, transmission, sharing, and use, and make them available to data users	Compliance with applicable laws on data sharing, transmission, and use is embedded in oversight and accountability functions, and is reflected in an organization's data strategy, which is published online	An organization's data transmission, sharing, and use activities fully comply with applicable laws and sometimes exceed legal standards
Data Ethics				
There is no awareness of the ethical questions that interoperable data might give rise to	There is a general understanding that joining up data may sometimes give rise to ethical questions, but it is unclear how	Active steps are taken to better understand the ethical impacts that joining up data might have and to understand how they might unintentionally cause harm Rudimentary ethical impact assessments are undertaken on an ad hoc basis	The types of ethical questions that joined-up data might give rise to are understood and appropriately categorized Appropriate steps are taken to minimize harm caused by a breach of ethical standards Ethical impact assessments are routinely undertaken	The risks of harm posed by joined-up data are well understood and ethical reviews are undertaken across the data life cycle to monitor issues and course correct as needed Ethical assessments are published transparently online An organization joins up data only once it has undertaken, and documented, a review of the potential risks of harm it might give rise to, and has taken appropriate steps to mitigate those harms
Procurement				
An organization is not aware of the impact of the procurement of technical and data solutions on interoperability	There is an emerging understanding of the need to join up data across procurement activities, but there is no coherent approach There is a heavy reliance on outside contractors to fill gaps but no coordination between units on how this is done, often resulting in incompatible data solutions being procured	Coordination across organizational units around procurement of technical solutions begins to materialize but is still not formalized Occasionally, units informally coordinate on the hiring of external contractors to ensure that new data systems are compatible with existing data infrastructure	The procurement of compatible and interoperable data systems across an organization is formalized and coordinated There is a common procurement policy across the organization that requires staff to consider interoperability issues when procuring new systems Reliance on external contractors is strategic and coordinated	An organization integrates the procurement of new interoperable software and data processing services into its data strategy and includes forward looking plans Units strategically plan and think through their common procurement needs and ensure that any new data system or service that is procured is both backwards compatible with existing infrastructure and meets likely future needs

ORGANIZATIONAL

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING

Links to Broader Data Ecosystems

There is no awareness of how data is used across a data ecosystem and the role of data interoperability in that

There is an emerging understanding that joined-up data across entities in a data ecosystem can give rise to both opportunities and challenges

An organization engages with other parts of the data ecosystem informally and in an ad hoc manner

An organization starts to attach pro forma licensing terms with provisions on data integration to data that it releases, transmits, or shares but does not monitor or engage with data users

An organization starts to document the data that it receives from other organizations

An organization starts to document and coordinate its engagements with other parts of the data ecosystem

An organization effectively categorizes its data and licenses it for use appropriately

An organization documents all data that is shared with it and has a general understanding of what it can and cannot do with it

An organization engages with other parts of the data ecosystem in a coordinated way, pursuant to its data strategy

There is a well-established and bespoke set of licenses that set out clear parameters for use, including integration depending on the category of data involved

An organization documents all data that is shared with it and has clear guidance and procedures in place that govern whether and how that data can be joined up with other data sets in its control

An organization makes engagement with other parts of the data ecosystem a strategic priority and has a well-coordinated approach with clear processes for joining up its data with external data

HUMAN

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
Data Stewardship				
No staff are assigned as data stewards	Joining up data forms part of certain staff members' function but it is not reflected in their job descriptions and is ad hoc	Staff in different units have recognized functions relating to data interoperability, but there is little or no central coordination	There is a coordinated staffing plan that reflects the various dimensions and roles relating to data stewardship, including interoperability, across the organization Units are coordinated and communicate with each other about what data they are joining up, how, and why	A strategically thought-through plan for data management is overseen by an organization's data governance council or committee and includes a clear plan for stewardship of data, including data interoperability functions Staff across the organization are aware of how data is used, joined up, and shared with other entities
Privacy and Confidentiality Preservation				
There is little to no understanding of the risks to privacy and the need to preserve confidentiality in interoperable data sets	There is emerging understanding of the risks posed to individuals or vulnerable groups if data is combined	There is awareness of applicable privacy and confidentiality related (international) law, normative principles, best practices, and guidance but they are not routinely considered or followed when an organization's data assets are integrated with other data or otherwise used	There is routine consideration of applicable law, principles, best practices, and guidance An organization undertakes privacy impact assessments before and during data-related projects and those assessments include considerations pertaining to the risks associated with data interoperability	The preservation of individual privacy and data confidentiality form part of an organization's legal and ethical review and are integrated across the data life cycle An organization adheres to the highest applicable standards of privacy and confidentiality preservation An organization integrates privacy and confidentiality preservation as part of its data strategy and explicitly provides guidance surrounding the risks of interoperable data, such as the **mosaic effect** An organization is forward-looking and cognizant of the potential privacy risks inherent to interoperable data posed by emerging technologies such as the Internet of Things, biometric ID verification, or general automated processes

HUMAN

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
Staff Knowledge and Skills				
Staff do not have the necessary knowledge or skills to join up data	Some staff have the knowledge and skills to join up data, but this is not reflected in their job descriptions and is tangential to their main functions	Knowledge and skills relevant to interoperability start to be recognized as part of job descriptions in some units, but the approach is fragmented	There is a coordinated approach to knowledge and skill strengthening across an organization that explicitly recognizes and addresses interoperability needs	The value of interoperability is recognized by numerous parts of the organization, including non-technical units Training courses relating to data governance issues, including interoperability, are available to all staff
Internal and External Communication				
There is no internally or externally coordinated communication reflecting the value of joined-up data	Examples of good practice and value generated as a result of interoperability emerge in an organization but are not communicated internally or externally	Staff and units start to share examples of good practice with each other, but this is not coordinated The value of interoperability starts to be understood by non-technical staff but is not yet communicated externally	Mechanisms to facilitate internal communication and sharing of best practices around interoperability form part of an organization's data strategy Cross-unit communication helps to translate best practices and examples of value generation for external audiences	An organization has a variety of coordinated internal communication channels open between units and staff, enabling the sharing of best practices and examples of value generation An organization is a champion of the value of joined-up data to data ecosystems and actively communicates its experiences and examples with others in compelling and effective ways, including through engagement with data journalists and storytellers
Adaptability				
Processes relating to staff functions and oversight of data interoperability are rigid and hard to change	There is emerging understanding of the value of adaptability in functions and oversight to data management generally, but no specific approach	Disparate units across an organization start to formally recognize the need to ensure that staff's functions and oversight of data systems are adaptable so as to ensure that value continues to be generated from their data assets	The value of empowering staff to be adaptable in how they use data, including in how they join it up with other data, is recognized by an organization and is reflected in its data strategy Staff have the authority to adapt their working processes and oversight of organizational data assets in ways that enhance its value, including by joining them up	An organization becomes a leader in adaptive management, and staff feel empowered and are confident in their ability to adapt their oversight of data systems as needed, including how they join up data, to maximize value

DATA

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING

Data and Metadata Modelling Capacity

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
There is little or no ability to model data or metadata	There is an emerging understanding of the value that data and metadata modeling can confer to data assets, but data modeling is not a priority for technical units	Disparate units across an organization recognize the value of data and metadata modeling, including its importance to data interoperability, and take steps to align their modeling techniques and start to coordinate their efforts	Technical units coordinate their approach to both data and metadata modeling and align efforts to consistently model data based on their organizational needs Internal needs are still prioritized over external groups, but data is modeled consistently	An organization routinely utilizes canonical data and metadata models that follow standardized patterns, making them reusable and conducive to data sharing The selection and application of canonical models is done through careful planning, including through engagement with data users and other entities in the data ecosystem

Data Organisation and Classification Capacity

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
An organization is unaware of the importance of data classification to interoperability and does not have a clear idea of its data assets	Units start to inventory their data Units are aware of the need for standardized data classification, but only use them on an ad hoc case-by-case basis	There are informal attempts between units to use common classifications, but these are not formalized or coordinated across all relevant units There is some, but not consistent, use of common classifications across the organization	There is a coordinated approach to the use of data classifications across the organization Units work together to identify the most appropriate classifications for their data and ensure that the data under their control is appropriately classified	The organization not only routinely and appropriately uses data classifications but also produces its own classifications to fill gaps and ensure consistency The organization engages actively with other entities in the data ecosystem to improve commonly used classification systems and establish new ones as needed The organization effectively communicates the value of consistent data classification for interoperability

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING

Data Access, Openness and Sharing

An organization has little or no knowledge of interoperability considerations when planning to responsibly manage data access, share data, or open it up for use	Disparate units across an organization are aware of interoperability considerations when planning, responsibly manage data access, share or publish data as open data, but this knowledge is not uniform or universally applied	There are coordinated efforts in technical units to ensure that data is accessible and shared responsibly, including relevant licensing permissions or limitations for future data integration and use Some data is made open on an organizational platform, but data sets are incomplete, not timely, or have not been quality assured	Data is shared responsibly in ways that protect any rights that third parties may have over it Data that is published openly is done so in machine readable formats under a clear open data license with terms of use, and has been stripped of attributes that may result in the re-identification of individuals or vulnerable groups Open data portals are accompanied by relevant contextual information and are visualized in ways that promote use by numerous audiences	An organization operates an effective data sharing policy that provides guidance on the various ways in which data sharing should take place, from publication under an open license, through to the use of data sharing or processing agreements Legal advice is available to staff wanting to share data that will be integrated with other data sets by third parties Open data is not just published in machine and human readable formats but is also made available as linked data through the semantic web There are feedback loops with key audience groups and the organization is responsive to user needs

Data Analytics and Automation

There is little to no awareness of how to enable interoperability between data sets to undertake data analytics or how to join up data to train algorithms (machine learning)	There is disparate understanding of the role of interoperability in undertaking automated data analytics across organizational units There is limited understanding of how interoperable data should be used to train algorithms	A coordinated approach between organizational units starts to emerge and some units start to produce scrubbed, quality assured, and consistent data sets that are available for integration and automated processing There is a coordinated effort to understand how data sets can be combined to train algorithms	Data analytics and machine learning functions are reflected in an organization's data strategy The relative benefits and risks of running automated analytics over interoperable data, or using it to train algorithms, are generally understood but there is not yet a consistent approach across an organization	An organization's data strategy includes forward looking plans for how data analytics tools can be responsibly applied to multiple, interoperable data sets in future There is a nuanced and well-established understanding of the relative benefits and risks of running automated analytics over interoperable data or using the data to train algorithms and appropriate risk and cost-benefit assessments are applied as needed An organization proactively engages with other entities in a data ecosystem to share its learnings and uses open-source analytics tools whenever possible to enable transparent scrutiny

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING

Data Protection

There is little to no understanding or awareness of the links between data interoperability and data protection techniques, including anonymization, pseudonymization, and encryption

There is some knowledge and understanding of the need to protect data that will be combined with other data, including through the use of appropriate pseudonymization, anonymization, and encryption techniques as needed, but this knowledge is not uniformly understood, and data protection techniques are not consistently applied

Disparate units routinely apply appropriate data protection techniques to their data sets before data integration, but there is little to no consistency in how those techniques are applied

There is some, but limited, understanding of the risks of re-identification inherent to interoperable data

Personal, sensitive, and sensitive group data is subject to appropriate protections before being integrated, shared, or processed through automated analytics tools

Risks of re-identification inherent to interoperable data are understood and are applied, but not routinely

All data is protected using the appropriate techniques and either responsibly archived or permanently deleted at the end of its intended life cycle

Access to sensitive data sets is monitored and documented to ensure accountability over data protection

Prior to integration, sharing, or processing through automated analytics, all data is assessed for risks of re-identification or other harms and is only used when there is a high degree of certainty that the data will remain safe following reuse

An organization helps to set standards for data protection within the broader data ecosystem and champions responsible data use

TECHNOLOGICAL

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING

Digital Infrastructure

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
An organization faces shortages of key infrastructure to store, manage, exchange, and process data, such as hardware and software components, a reliable electricity supply, or Internet connectivity	There is adequate access to key infrastructure components but there is a shortage of organization-specific data storage and content management solutions, resulting in non-standardized and non-aligned data management systems	All appropriate staff members have access to adequate hardware and software tools, as well as network connectivity There are secure servers and data repositories, but they are used inconsistently by staff and organizational units; there is little oversight of digital infrastructure	All appropriate staff members are aware of, and trained in, how to use an organization's data management and processing systems Secure servers and data repositories are routinely used by staff members and oversight of digital infrastructure is part of an organization's data strategy	An organization's data strategy includes provisions for the maintenance, regular review, and upgrading of its digital infrastructure, and budget lines are set aside for this purpose Data policies and standards on procurement, data sharing, and infrastructure oversight are aligned An organization is forward thinking in its approach to digital infrastructure and actively strategizes and plans on how it can make best use of emerging technology to improve the interoperability of its data systems

Cybersecurity and Incident Response

UNDEFINED	EMERGING	LEARNING	BUILDING	CONSOLIDATING
There is little to no awareness of the risks of cyberattacks or other breaches to an organization's data systems, including the specific risks associated with potentially reusable, interoperable data No data breach protocol or policy is in place	Disparate staff and units across an organization have awareness or show concern about the risks posed to their reusable data by a cyberattack or other data breach Champions emerge who push for a data breach protocol or policy	A data breach protocol is drafted, but risks associated with the potential reuse of stolen interoperable data remain vague and there is inconsistent understanding and application of the policy	A clear data breach policy setting out sequential steps and responsibilities is established Staff receive training on what they should do in the event of a data breach and are taught about the risks associated with the reuse of interoperable stolen data	An organization is able to deal with data breaches swiftly and effectively, and takes active steps to ensure that its technological infrastructure is as secure as possible The data breach policy is regularly reviewed and updated, and explicitly covers risks associated with interoperable data reuse Appropriate staff are routinely trained on how to respond to a data breach